



TITLE:

市民参加型史料研究のためのデジタル人文学基盤の構築(
Abstract_要旨)

AUTHOR(S):

橋本, 雄太

CITATION:

橋本, 雄太. 市民参加型史料研究のためのデジタル人文学基盤の構築. 京都大学, 2018, 博士(文学)

ISSUE DATE:

2018-07-23

URL:

<https://doi.org/10.14989/doctor.r13199>

RIGHT:

京都大学	博士（文学）	氏名	橋本 雄太
論文題目	市民参加型史料研究のためのデジタル人文学基盤の構築		
<p>（論文内容の要旨）</p> <p>本論文は、デジタル人文学（Digital Humanities）における市民参加型研究の手法を、前近代日本語史料を対象に実現する試みである。近年、さまざまな学術分野において、インターネットを介して多数の市民の協力を募り、大規模なデータ処理やデータ収集を実現するプロジェクトが活発に実施されている。これらは「クラウドソーシング」や「市民科学」といった呼称で呼ばれるが、本研究では「市民参加型研究」と総称している。人文学領域では、人文学と情報学の融合分野であるデジタル人文学分野を中心として、文献資料の翻刻（文字起こし）に市民参加型研究の手法が頻繁に用いられている。たとえば英国ユニバーシティ・カレッジ・ロンドンが運営する Transcribe Bentham は、6 万ページに及ぶジェレミ・ベンサムの手稿を、インターネットを介して参加する多数の市民ボランティアの手により翻刻するプロジェクトであり、すでに 1 万 9 千ページが同プロジェクトを通じて翻刻されている。一方、古文書、古記録、古典籍など、日本国内には江戸時代以前から伝来する膨大な点数の文献史料が未翻刻のまま残されているにも関わらず、歴史資料に対する市民参加型研究が実現した例はこれまでになかった。本研究は、災害史料という限定されたジャンルではあるものの、前近代日本語史料を対象とする市民参加型研究を実現した、最初の研究である。本論文の構成は以下の通りである。</p> <p>第 1 章は、研究の歴史的背景の説明である。インターネットは高速かつ安価な通信手段を提供しただけでなく、人々が自発的にコミュニティを形成する「オンライン参加」を可能にした。「オンライン参加」は、Web 2.0 の時代に技術的知識を持たない一般の人々にも広まった。この様な背景のもと、Wikipedia、Project Gutenbergなどの大規模参加型プロジェクトが生まれ、さらには、オンライン参加の商用利用も出現した。そして、2000 年代後半からは、学術領域においても、オンライン参加の利用が始まった。デジタル人文学の分野では、特に、歴史文献史料を対象とした市民参加型研究が活発であり、大きな成果を収めている。しかしながら、前近代日本語史料を対象とした市民参加型研究が成功したことはこれまでになかった。これは、(1)文字解読（くずし字解読）の問題、(2)参加者の動員の問題、(3)参加者の動機付けの問題、が障害となっていたためである。</p> <p>第 2 章では、本研究の実現に大きな役割を果たした、『くずし字学習支援アプリ KuLA』について述べる。KuLAは、大阪大学文学研究科が中心となって開発した、「くずし字」解読能力習得支援アプリである。筆者はこのプロジェクトに特任研究員として参加し、設計と実装の全工程を担当した。KuLA は、公開後 1 ヶ月で 1 万ダウンロードを達成するなど、学術分野で開発したアプリケーションとしては異例のダウンロー</p>			

ド数を記録した。またKuLA のコミュニティ機構を用いる活発なオンライン参加行動が観察された。これにより、くずし字解読においても、オンライン参加をベースにした大規模プロジェクトの実施が可能であること、そのために「学習」がきわめて有用な手段として機能するという二つの知見が得られた。

第3章では、本研究の中心的成果である歴史災害史料の市民参加型翻刻プロジェクト『みんなで翻刻』の設計と実装、またシステム公開後の成果について述べる。過去に発生した地震の記録は、地震発生のメカニズム究明や将来的な地震予測のために不可欠な研究データである。江戸時代以前に発生した地震の情報は、古文書や古記録から入手する必要がある。しかし莫大な災害史料を、少数の研究者の手によって翻刻するのは困難で、翻刻作業の能率向上が課題だった。そこで、インターネットを介して多数の市民の協力を募り、地震史料の翻刻を一挙に進める計画が、筆者も属する京都大学古地震研究会において立てられた。それが『みんなで翻刻』である。『みんなで翻刻』では、KuLAの知見や、先行する外国語学習サービスをヒントに、「翻刻作業を通じた学習」と「参加者間の協働」の実現を基本方針とした。この方針のもと、翻刻作業の参加者が互いの翻刻文を共有し、難読箇所や未読箇所の読み方を教え合うことを可能にする相互添削システムが実装された。2017 年 1 月のシステム公開後、予想を遥かに上回るペースで翻刻作業が進行した。プロジェクト開始当初の目標であった、東京大学地震研究所が所蔵する災害史料コレクション「石本文庫」の収録史料 114 点（画像 3,193 枚）の全文翻刻は、システム公開から 5 ヶ月弱の期間で完了した。公開から 13ヶ月が経過した 2018 年 2 月時点では、3,800 人の参加登録者によって、石本文庫を含む 412 点の史料（画像 5,491 枚）が全文翻刻されており、入力された翻刻文は 400 万文字を超えている。また日本史研究者の協力のもと、翻刻文の正確性を検証した所、翻刻文中に含まれる誤刻箇所や未読箇所などは、100 文字あたり平均1.5 文字だった。この品質は、学術出版される史料集の品質には及ばないものの、史料の内容把握や全文検索には十分な水準であり、『みんなで翻刻』は、前近代日本語史料を対象にした市民参加型研究の最初の成功例と言える。

第4章は、『みんなで翻刻』の成功の要因分析である。市民参加型の翻刻プロジェクトには、大きな成果を挙げたものもあれば、ほとんど成果の挙げられなかったものもある。多種多様なプロジェクトの明暗の要因をはっきりと特定することは難しいが、具体的な成功例と失敗例の比較により、「マスメディアによる報道の有無」と「参加者の動機付けの有無」の2つが、プロジェクトの成否に大きな影響を及ぼしたらしいことを指摘した。続いて、(1)文字解読の問題、(2)参加者の動員の問題、(3)参加者の動機付けの問題といった諸課題が、『みんなで翻刻』においてどのように解消されているか、参加者を対象にしたアンケート調査の結果や、Webサイトのアクセス集計といったデータの分析を通じて明らかにした。まず文字解読の問題については、大学で日本史や国文学を専攻した、くずし字解読の経験者が多数プロジェクトに参画したことに加えて、参加者が相互にコミュニケーションを取りながら正確な翻刻文を作り上げ

る相互添削システムが有効に機能したことが、大規模な史料翻刻を可能にしたと見られる。次に参加者の動員については、マスメディア報道が後押しとなったことは事実であるが、一方で参加者を対象としたアンケートの回答者の7割が KuLA のユーザーであると回答するなど、KuLA が『みんなで翻刻』へ参加者を誘導する「導線」として機能していた可能性も示唆された。参加者の動機付けについては、アンケート調査の回答から、「翻刻作業そのものの楽しさ」に次いで、「参加者間の協働を通じた学習」が大きな役割を果たしていることが読み取られた。以上の分析をふまえて、『みんなで翻刻』の成果は、複数の学習者が互いに協働することによって学習を進める教育手法である「協調学習 (collaborative learning)」の結果として解釈できることを指摘し、さらに協調学習を基礎として、『みんなで翻刻』の設計を、他のプロジェクトにおいても再利用可能な設計パターンとして定式化する試みをおこなった。

第5章では、災害史料の市民参加型翻刻プロジェクトである『みんなで翻刻』の対象を拡張し、前近代日本語史料一般に適用する意義と可能性について検討する。前近代日本語史料に市民参加型翻刻の手法を適用する上でまず問題となるのは、専門的訓練を受けていない市民が活動の主体を担う『みんなで翻刻』のようなプロジェクトに、そもそも歴史学の立場から学術的意義を認めうるか、という問題である。『みんなで翻刻』の成果物は、専門家によるレビューを受けておらず、厳密な学術出版のプロセスを経た刊行史料集と同等の研究資料として扱うには問題がある。しかしながら『みんなで翻刻』は1年あまりの期間に400万文字以上の史料翻刻を実現しており、学術出版のレベルには劣るものの内容把握や全文検索には問題ない品質を達成している。したがって、「学術分野の翻刻としては品質面で問題があるが、史料画像を参照することで検証が可能であり、大まかな内容把握や情報検索、あるいは厳密な翻刻のドラフトとして利用するには有用である」というコンセンサスが専門家と市民の間に成立する限りにおいては、市民参加型研究の成果に学術的な意義を認めることは可能である。以上の議論を通じて市民参加型研究の意義を確認した上で、複数デジタルアーカイブと連携し日本語史料一般の市民参加型翻刻を可能にする拡張版『みんなで翻刻』の構想を素描した。拡張版『みんなで翻刻』は、複数プロジェクトの同時稼働と IIF API を介した複数デジタルアーカイブとの連携に対応することで、一般の歴史資料の市民参加型翻刻を実現可能にする。さらに、SMART-GS のような研究者向けのシステムを取り込むことで、市民参加によって得られた成果物を、研究者や学芸員といった専門家が学術研究に利用することを可能にするものである。この拡張版『みんなで翻刻』により、ごく近い将来に、東寺百合文書や国文学研究資料館新日本古典籍データベース、国立国会図書館デジタルコレクションなど、大規模な歴史資料群の市民参加型翻刻が可能となるはずである。

(論文審査の結果の要旨)

ITを利用する人文学であるデジタル・ヒューマニティーズ Digital Humanities の近年の進歩には目を見張るものがある。本論文は、その一分野である Crowd Transcription (以下、クラウド翻刻と書く)における研究と実践を報告するものである。

従来、歴史資料の翻刻は、歴史家や歴史家のチームにより行われるものであった。たとえば、本研究科の名誉教授永井和氏をリーダーとする歴史家のチームによる倉富勇三郎日記の翻刻プロジェクトは、その典型である。

しかし、膨大な歴史資料を、専門家のみで翻刻することには、量的限界がある。そのため、市民のボランティア参加による歴史資料の翻刻プロジェクトが数多く実施されている。たとえば、哲学者J. ベンサムの手稿を翻刻する、英国ユニバーシティ・カレッジ・ロンドンの Transcribe Bentham や、米国スミソニアン博物館所蔵の手稿を翻刻する Smithsonian Transcription Center などが、代表的成功例として知られており、前者では約2万枚、後者では30万枚をこえる史料が既に翻刻されている。

これらの事例では、史料の画像が、WEBページ上に表示され、それをボランティアが翻刻して、同じページのテキストボックスに打ち込む。この様な、ネット、WEB上で、人々 (Crowd) によりなされるプロジェクトを、クラウドソーシング、クラウド・コンピューティングなどと呼ぶため、これらはクラウド翻刻と呼ばれる。

このような海外、主に英語圏のクラウド翻刻の成功事例の一方で、我が国では、手書き史料のクラウド翻刻の成功例が全く存在しなかった。そればかりか、日本デジタル・ヒューマニティ学会による、国会図書館所蔵の活字史料のクラウド翻刻さえ、公開したものの数枚の翻刻しかなされないという状況であった。この閉塞状態に風穴を開けたのが、論者が開発し運営している「みんなで翻刻」である。

この「みんなで翻刻」というクラウド翻刻システムは、京都大学古地震研究会の活動の一環として、2017年2月に公開されたもので、東大地震研が所蔵する、代官所の記録などの、主に江戸時代以前の地震、噴火などの災害に関連する古文書を、ボランティアが翻刻するというものである。この「みんなで翻刻」は、公開以来、1年と少しで、6千枚以上の翻刻を完了しており、これは2014年に公開された Smithsonian Transcription Center には、及ばないものの、2010年に公開された Transcribe Bentham での1年あたりの翻刻枚数を凌ぐものであり、我が国のデジタル・ヒューマニティーズにおける大きなブレイク・スルーと言える。

論者は、本論文の第1章で、世界と我が国のクラウド翻刻の状況を、インターネットの黎明期からのクラウドソーシングの歴史から説き起こして説明する。そして、我が国におけるクラウド翻刻、特に明治より前の古文書のクラウド翻刻の困難さを、(1)文字解読(くずし字解読)の問題、(2)参加者の動員の問題、(3)参加者の動機付けの問題、の三つに求める。

そして、論者は、第2章で、これらの問題を克服する方法の考案に大きな助けとなった、大阪大学文学研究科が中心となって作成された、くずし字解読学習支援スマート

フォン・アプリ KuLA の開発経験を述べる。このアプリのコンテンツは、阪大などにより開発されたが、アプリそのものと、ユーザー間の相互支援の仕組みなどは、特任研究員として雇用された論者が考案・開発したものである。論者は、この開発経験と、KuLAのユーザーたちの行動を観察したことが、ボランティアに、くずし字解読の学習の機会を提供し、また、ボランティア間の学習相互支援を可能にした、くずし字のクラウド翻刻システムである「みんなで翻刻」につながったとしている。

第3章は、そのシステム構成など、論者が独力で案出し、また、構築した「みんなで翻刻」の技術的説明や、論者が、クラウド翻刻の成功のための重要要因とする、参加者の動員のために行った広報活動などが説明された上で、その後の大きな成功の経緯が語られる。

続く第4章では、この成功の要因が、ボランティアに対して行ったアンケート調査などを元に分析され、論者は「みんなで翻刻」が、教育学でいう協調学習の場になっていたのが成功の大きな要因だと結論する。そして、最終章である第5章では、この協調学習という成功要因の理論づけにもとづき、我が国におけるクラウド翻刻を成功に導くためのモデルが提唱され、さらに、「みんなで翻刻」が持つ学術的意味と、将来に向けての課題・計画などが議論されて、本論文は終わる。

繰り返しとなるが、英語圏などでは、すでに実用段階に達している手書き史料のクラウド翻刻を、我が国において初めて成功させただけでなく、論者は、それをシステムの設計、実現、運営においては、ほぼ一人で達成している。たとえば、Transcribe Bentham が、58万ポンドの予算と多くの専属スタッフを擁する巨大プロジェクトであることを考えれば、論者があげた研究成果は驚くべきものだといえる。

ただし、このような実践的研究の成功の一方で、本論文にはいくつかの問題点があることも確かである。審査委員からは、協調学習モデルの意義への疑問、アンケートの分析方法に若干の問題があること、さらには、学習の場を提供したことが成功の要因とする議論において、翻刻の大半が比較的少数のボランティアによって行われているという事実が十分考慮されていない、などの問題点が指摘された。

しかし、これらの問題点はあるものの、本論文が我が国の古文書のクラウド翻刻が「みんなで翻刻」によって可能となった、その仕組みの一端を解明したという事実は揺るがない。また、これらの問題点は、今後の「みんなで翻刻」のさらなる運用や、類似のクラウド翻刻の実施を通して、解決されていくことであろう。論者の一層の精進を期待したいところである。

以上、審査したところにより、本論文は博士（文学）の学位論文として価値あるものと認められる。2018年5月16日、調査委員4名が論文内容とそれに関連した事柄について口頭試問を行った結果、合格と認めた。